Method for Improving Results in an HMM-based Segmentation System by Incorporating External Knowledge

Abstract of the Invention

A Hidden Markov model is used to segment a data sequence. To reduce the potential for error that may result from the Markov assumption, the Viterbi dynamic programming algorithm is modified to apply a multiplicative factor if a particular set of states is re-entered. As a result, structural domain knowledge is incorporated into the algorithm by expanding the state space in the dynamic programming recurrence. In a specific example of segmenting resumes, the factor is used to reward or penalize (even require or prohibit) a segmentation of the resume that results in the re-entry into a section such as Experience or Contact Information. The method may be used to impose global constraints in the processing of an input sequence or to impose constraints to local sub-sequences.